



Estimation of a Change Point in a Hazard Function Based on Censored Data¹

IRÈNE GIJBELS

Université Catholique de Louvain, Belgium

gijbels@stat.ucl.ac.be

ÜLKÜ GÜRLER

Bilkent University, Turkey

ulku@bilkent.edu.tr

Received April 10, 2002; Revised May 1, 2003; Accepted June 5, 2003

Abstract. The hazard function plays an important role in reliability or survival studies since it describes the instantaneous risk of failure of items at a time point, given that they have not failed before. In some real life applications, abrupt changes in the hazard function are observed due to overhauls, major operations or specific maintenance activities. In such situations it is of interest to detect the location where such a change occurs and estimate the size of the change. In this paper we consider the problem of estimating a single change point in a piecewise constant hazard function when the observed variables are subject to random censoring. We suggest an estimation procedure that is based on certain structural properties and on least squares ideas. A simulation study is carried out to compare the performance of this estimator with two estimators available in the literature: an estimator based on a functional of the Nelson-Aalen estimator and a maximum likelihood estimator. The proposed least squares estimator turns out to be less biased than the other two estimators, but has a larger variance. We illustrate the estimation method on some real data sets.

Keywords: change point hazard model, cumulative hazard function, least squares estimation, random censoring

1. Introduction

Suppose we are interested in the analysis of a quantity that can generally be expressed as the time to some event, such as the time to failure of a component, the time from infection with HIV until sero-conversion in AIDS epidemic or the survival time after a certain operation. We will refer to this quantity as the lifetime variable and denote it by X , a nonnegative random variable. Also, the termination of the lifetime will be referred to as ‘failure’. In the analysis of data related to the variable mentioned above, the hazard rate $\lambda(x)$, also defined as the instantaneous failure rate function plays an important role since it describes the immediate risk of failure of an individual (or an item) at time x , given that it has not failed up to that time. In practical situations abrupt changes in the hazard function at unknown points can be encountered due to for example a new treatment in survival analysis or a major maintenance action in reliability.

In this paper we consider estimation methods for hazard functions with a jump discontinuity, when the observed data is subject to right censoring. In medical research, one often has to deal with right censored data for the survival times of the patients due to drop outs from the study or to the termination of the observation period. To be more precise, suppose the interest is in the variable X but instead of observing a random

independent and identically distributed (i.i.d.) sample of X , we observe the random sample (T_i, δ_i) , $i = 1, 2, \dots, n$, where $T_i = \min(X_i, C_i)$ and $\delta_i = 1$ if $X_i \leq C_i$ and zero otherwise. Here the random variable C is the censoring variable which is assumed to be independent of X . Due to this censoring effect, the methods available for the analysis of i.i.d. data are clearly not appropriate in this set-up and other methods are needed. For continuous hazard rate functions, parametric and nonparametric estimation methods are available in the literature. For hazard functions with discontinuities, the literature is not so rich as will be discussed briefly below.

Suppose that the hazard function of X is constant with value β until τ , at which point it makes a jump of size θ and stays constant (of value $\beta + \theta$) thereafter. Here the interest is in estimating the unknown parameter vector (τ, β, θ) . This problem belongs to the class of so-called change point problems. The literature on change point problems in the censored data situation is rather small. There are mainly three different approaches for inference under this model. A first approach is based on the use of maximum likelihood estimation methods and is considered by Matthews and Farewell (1982), Nguyen, Rogers and Walker (1984), and Loader (1991). Another issue is testing for a constant hazard rate versus an alternative involving a change-point as in the model described above. This problem is considered by Matthews, Farewell and Pyke (1985) using score statistics and later by Henderson (1990) and Loader (1991). Finally, Chang, Chen and Hsiung (1994) exploit the structural properties under the above simple change point model. Their estimation procedure relies on a certain functional of the estimated cumulative hazard function. A nice overview of the existing literature can be found in Müller and Wang (1994). See also Müller and Wang (1990).

In this paper we propose an alternative approach for estimation of τ , the location of the jump point. This estimator is then used to obtain estimators for β and θ . Our approach relies on certain structural properties of the cumulative hazard function. With the estimated cumulative hazard we then fit an appropriate regression model to synthetically obtained data.

The paper is organized as follows. In the remainder of the introduction we describe two examples which motivate the problem studied in this paper. In Section 2 we describe the change point model and discuss estimation methods for it, including a new estimation approach based on least squares ideas. In Section 3 we analyze the data described in Section 1.1, using the change point model and the discussed methods. A simulation study, presented in Section 4, reveals the specific merits of each of the three estimation methods.

1.1. Data Examples

We now describe two data sets which we will analyze using the change point model approach.

Lymphoma Data: The available data is the one on non-Hodgkin's lymphoma described and analyzed in Matthews, Farewell and Pyke (1985), which consists of the survival times (in months), defined as the time from diagnosis to death, of 31 individuals with advanced non-Hodgkin's lymphoma with clinical symptoms. Of these, 11 are censored because the patients were alive at the last time of follow-up, including the largest observation which is treated as an uncensored data in the analysis. The data are 2.5, 4.1, 4.6, 6.4, 6.7, 7.4, 7.6,

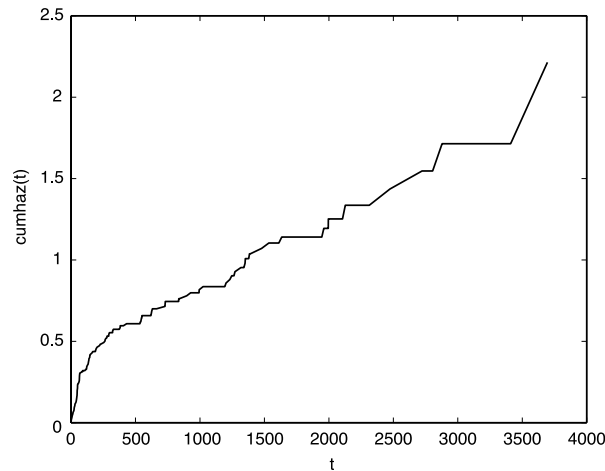


Figure 1. Estimated cumulative hazard for the Stanford heart transplant data.

7.7, 7.8, 8.8, 13.3, 13.4, 18.3, 19.7, 21.9, 24.7, 27.5, 29.7, 30.1*, 32.9, 33.5, 35.4*, 37.7*, 40.9*, 42.6*, 45.4*, 48.5*, 48.9*, 60.4*, 64.4*, 66.4*, where an (*) indicates a censored observation.

Matthews, Farewell and Pyke (1985) tested the null hypothesis of a constant hazard rate against the alternative of failure rates involving a single change point and reported a strong evidence against constant hazard rate in favor of the alternative. In Section 3, we analyze this data set, using a simple change point model, and estimate the location and size of the assumed change point as well as the hazard function. From our analysis we can conclude that the risk of dying from the disease is approximately 0.030 during the first eighteen months following the diagnosis and that after this period the risk decreases slightly and drops to about 0.019.

Heart Transplant Data: The second data set we consider is the well known Stanford heart transplant data, which is previously analyzed by several authors, including Miller and Halpern (1982), and Cox and Oakes (1984). We use the data set as given in Miller and Halpern (1982). These data consist of the survival times of 184 patients having received a heart transplant. Among these observations 113 were censored. The survival times (in days) start from 0.0 and the largest observed survival time is 3695, which is censored, whereas the largest uncensored survival time is 2878. There are many ties in the sample. The empirical cumulative hazard function is depicted in Figure 1 and a kernel smoothed nonparametric estimator of the hazard rate function is given in Figure 2 (using a bandwidth equal to 70 days). Looking at Figure 2 we see that the estimated hazard function shows relatively high values short after the transplantation. The hazard function then drops quite rapidly to a lower level. The analysis of this data set based on the change point model, presented in Section 3, revealed that the instantaneous risk of failure for patients who received heart transplantation changes significantly after a critical period of

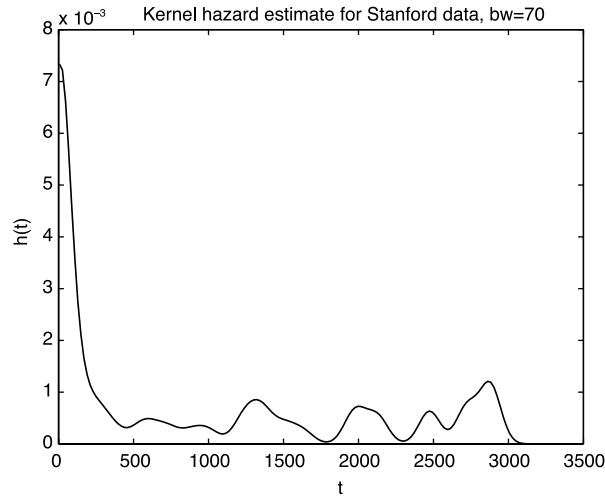


Figure 2. Kernel estimator of the hazard rate function for the Stanford heart transplant data.

about 70 to 80 days. During the first 80 days after the transplantation, the instantaneous risk of failure is approximately 0.0040, whereas after this period the risk drops to approximately 0.0004, one tenth of the previous level.

2. The Model and Estimation Methods

Consider a change point hazard model where the hazard function $\lambda(x)$ of the random variable X is modeled as:

$$\lambda(x) = \beta + \theta I_{\{\tau < x\}}, \quad (2.1)$$

where β and θ are such that $\beta > 0$ and $\beta + \theta > 0$. Under this model the cumulative hazard function is given by

$$\Lambda(x) = \int_0^x \lambda(t) dt = \beta x + \theta(x - \tau) I_{\{\tau < x\}}.$$

We first briefly present the maximum likelihood estimators and the estimation procedure proposed by Chang, Chen and Hsiung (1994). The alternative estimation procedure, based on a least squares idea will be explained in Section 2.3.

For each of the procedures we will focus on the estimation of the unknown location point τ , but of interest is also to estimate β and θ in model (2.1), i.e. the value of the hazard rate before the jump point and the size of the jump in τ , respectively. Once an estimator for

τ is available estimates for β and θ are simply obtained by classical maximum likelihood estimators. These are described in (2.3), where we replace τ by its estimated value. As can be seen from these expressions, the estimation of β relies on that of τ . The estimator for θ will be influenced by the estimates of both τ and β . In our simulation study in Section 4 we will evaluate the performances of the estimators for τ , β and θ .

2.1. Maximum Likelihood Estimation

The density function of the variable X can be obtained via its hazard function modeled in (2.1):

$$f(x) = \lambda(x) \exp\left\{-\int_0^x \lambda(t) dt\right\} = \begin{cases} 0 & \text{if } x < 0 \\ \beta \exp(-\beta x) & \text{if } 0 \leq x \leq \tau \\ (\beta + \theta) \exp\{-\beta x - \theta(x - \tau)\} & \text{if } x > \tau. \end{cases}$$

Let $X(t)$ denote the number of deaths observed up to time t :

$$X(t) = \sum_{i=1}^n I(T_i \leq t) \delta_i,$$

and let n_u be the number of uncensored observations. Consider the case of non-informative censoring (meaning that the censoring distribution is independent of the parameters τ , β and θ). Then, for model (2.1) with the above expression for the density f , the log-likelihood function can be written as (see e.g., Cox and Oakes, 1984; Loader, 1991)

$$\begin{aligned} \log L(\beta, \theta, \tau) &= X(\tau) \log \beta + (n_u - X(\tau)) \log(\beta + \theta) \\ &\quad - \beta \sum_{i=1}^n \min(T_i, \tau) - (\beta + \theta) \sum_{i=1}^n (T_i - \tau) I(T_i > \tau). \end{aligned} \quad (2.2)$$

For fixed τ the maximum likelihood estimates of the parameters β and θ are given by

$$\hat{\beta} = \frac{X(\tau)}{\sum_{i=1}^n \min(T_i, \tau)} \quad \text{and} \quad \hat{\theta} = \frac{n_u - X(\tau)}{\sum_{i=1}^n (T_i - \tau) I(T_i > \tau)} - \hat{\beta}. \quad (2.3)$$

Substitution of these estimates for β and θ into the log-likelihood function in (2.2) leads to the objective function

$$\log L(\tau) = -n_u + X(\tau) \cdot \log \frac{X(\tau)}{\sum_{i=1}^n \min(T_i, \tau)} + (n_u - X(\tau)) \log \frac{n_u - X(\tau)}{\sum_{i=1}^n (T_i - \tau) I(T_i > \tau)}.$$

Now, maximizing this objective function over $\tau \in [\tau_0, \tau_1]$, with $0 \leq \tau_0 < \tau_1 < \infty$ and τ_1 strictly smaller than the largest uncensored observation, leads to the maximum likelihood

estimator of τ , denoted by $\hat{\tau}_{\text{MLE}}$. Distribution theory for $\hat{\tau}_{\text{MLE}}$ has been studied by Yao (1986). He showed that $\hat{\tau}_{\text{MLE}} - \tau = O_P(n^{-1})$ under certain conditions.

Note that this maximum likelihood estimation of τ involves the choice of an interval $[\tau_0, \tau_1]$. This interval can be chosen quite wide in practice, with the only important requirement that the largest uncensored observation should be falling above τ_1 .

2.2. Estimation Method of Chang, Chen and Hsiung

Chang, Chen and Hsiung (1994) assume that the unknown location point τ belongs to a certain known interval $[\tau_0, \tau_1]$ with $0 < \tau_0 \leq \tau \leq \tau_1 < \infty$. Their method is based on the following functional of the cumulative hazard function $\Lambda(\cdot)$:

$$Y_{\text{CCH}}(t) = \left[\frac{\Lambda(T) - \Lambda(t)}{T - t} - \frac{\Lambda(t) - \Lambda(0)}{t} \right] g(t(T - t)), \quad (2.4)$$

for $0 < t < T$, where $T > \tau_1$ and $g(x) = x^p$, $0 \leq p \leq 1$. When $\theta > 0$, it can be shown that $Y_{\text{CCH}}(t)$ is increasing on the interval $[0, \tau]$ and decreasing on the interval $[\tau, T]$, hence the maximum of $Y_{\text{CCH}}(t)$ occurs at τ . This motivated Chang, Chen and Hsiung (1994) to consider

$$\hat{\tau}_{\text{CCH}} = \inf\{t \in [\tau_0, \tau_1] : Y_n(t\pm) = \sup_{u \in [\tau_0, \tau_1]} Y_n(u)\}, \quad (2.5)$$

with $Y_n(t\pm)$ denoting the right-hand and left-hand limits, and where $Y_n(t)$ is the empirical version of (2.4) obtained by replacing the unknown cumulative hazard function by the Nelson-Aalen type estimator. Let $Z_1 < Z_2 < \dots < Z_n$ be the order statistics corresponding to T_1, T_2, \dots, T_n . Then the empirical cumulative hazard is given by (see Aalen, 1978):

$$\Lambda_n(t) = \sum_{i: Z_i \leq t} \frac{\delta_i}{n - i + 1}. \quad (2.6)$$

The estimator $Y_n(t)$ of $Y_{\text{CCH}}(t)$ is obtained by replacing $\Lambda(t)$ in (2.4) by $\Lambda_n(t)$. The estimator $\hat{\tau}_{\text{CCH}}$ suggested by CCH is simply the point where the maximum of $Y_n(t)$ occurs within the interval $[\tau_0, \tau_1]$. Observing that $Y_n(t)$ is monotone on each interval $[Z_i, Z_{i+1})$, $\hat{\tau}_{\text{CCH}}$ equals to either one of the uncensored times Z_i , $i = 1, 2, \dots, n$ or one of the end points τ_0 or τ_1 .

When $\theta < 0$, the function $Y_{\text{CCH}}(t)$ in (2.4) is decreasing on $[0, \tau]$ and increasing on $[\tau, T]$ and an estimator similar to $\hat{\tau}$ in (2.5) can be obtained:

$$\hat{\tau}_{\text{CCH}} = \inf\{t \in [\tau_0, \tau_1] : Y_n(t\pm) = \inf_{u \in [\tau_0, \tau_1]} Y_n(u)\}.$$

Chang, Chen and Hsiung (1994) show that $\hat{\tau}_{\text{CCH}}$ is a consistent estimator of τ , with $\hat{\tau}_{\text{CCH}} - \tau = O_P(n^{-1})$.

Note that the estimation procedure of Chang, Chen and Hsiung (1994) involves the choice of the interval $[\tau_0, \tau_1]$ and the choice of the truncation point T . Although we use the same notation for the interval over which the candidate estimate for τ is searched for, the choice of this interval is far more restrictive in practice than in the maximum likelihood procedure, as will be seen from the simulations.

2.3. The Least Squares Estimator

The estimation procedure of the previous section is based on the particular functional $Y_{\text{CCH}}(t)$ as defined in (2.4). This functional is basically considering the differences of slopes of the cumulative hazard function. We base our least squares estimation procedure on the following function

$$Y(x) = \frac{\Lambda(x)}{x},$$

which is the slope of the line joining the points $(0, \Lambda(0))$ and $(x, \Lambda(x))$ lying on the curve of the cumulative hazard function $\Lambda(\cdot)$. In a more intuitive sense, $Y(x)$ can be viewed as the *average hazard* at time x .

Under model (2.1) we have

$$Y(x) = \beta + \theta \left(1 - \frac{\tau}{x}\right) I_{\{\tau < x\}}. \quad (2.7)$$

Hence, under the assumed model, the function $Y(\cdot)$ has a simple structure: it remains constant up to time τ and from τ on starts increasing (in case $\theta > 0$) or decreasing (in case $\theta < 0$) as a function of $1/x$.

Denote by $Y_n(x)$ the empirical version of $Y(x)$ obtained by replacing the unknown cumulative hazard function by the Nelson-Aalen estimator given in (2.6). The least squares procedure now consists of fitting, via least squares, a constant line up to τ and from τ on a function of the form $\beta + \theta(1 - \tau/x)$, through the ‘data points’ $(x_i, Y_n(x_i))$, $i = 1, \dots, n_g$, where x_i ’s are grid points to be chosen. The splitting point τ which would give the best least squares fit would give the estimator for τ .

We now describe in detail the least squares estimation procedure. Let τ_0, τ_1 be the wide interval over which the estimated value of τ is searched for. Of course this interval should be such that it contains the true value of τ . We start by considering a fixed partition of this interval via n_g grid points denoted by $\tau_0 = x_1 < x_2 < \dots < x_{n_g} = \tau_1$. Since our method consists of carrying out a least squares fit through the points $(x_i, Y_n(x_i))$, we need to make sure that we do have sufficient ‘data points’ for the fitting. To the left-hand side we fit a constant line, and hence this can be carried out even when there is only one ‘data point’. The fitting of the function to the right-hand side typically requires more ‘data points’. In order to take care of this we include the option of carrying out the least squares fit on a subset of the initial grid points x_1, x_2, \dots, x_{n_g} . Recall that n_u denotes the number of uncensored observations in the sample, and denote by $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(m_u)}$ the ordered

uncensored observations. We will search for the estimator of the location point τ over a possible reduced set of fixed grid points, such that we have a guaranteed portion of the actual data falling to the left of the first grid point in this reduced set, and to the right of the last grid point in the reduced set. Let α_1, α_2 be these proportions in the left and right tail respectively. The choice $\alpha_1 = 0$ is completely justified, but for α_2 one preferably would take a small positive value. Define $m_\ell = [\alpha_1 \times n_u]$ and $m_u = [(1 - \alpha_2) \times n_u]$ and get $T_{(m_\ell)}$ and $T_{(m_u)}$. We then consider

$$\begin{aligned} x_\ell &= \text{the smallest grid point in the initial grid that exceeds } T_{(m_\ell)} \\ &= \min\{x_j : j = 1, \dots, n_g \text{ such that } x_j > T_{(m_\ell)}\} \\ x_u &= \text{the largest grid point in the initial grid that does not exceed } T_{(m_u)} \\ &= \max\{x_j : j = 1, \dots, n_g \text{ such that } x_j < T_{(m_u)}\} \end{aligned}$$

and search the candidate for τ , via least squares fitting, on the interval $I_{\alpha_1, \alpha_2} = [x_\ell, x_u]$. So, the set of grid points over which we search the candidate estimator changes with the sample, but the grid points itself are fixed from the start. In the simulation study we illustrate the performance of the least squares procedure for various values of α_1 and α_2 . That is, for each point $x_i \in I_{\alpha_1, \alpha_2}$ we carry out the following least squares fit:

minimize

$$L(x_i, \beta, \theta) = \sum_{j=1}^i [Y_n(x_j) - \beta]^2 + \sum_{j=i+1}^{n_g} [Y_n(x_j) - \beta - \theta(1 - x_{ij})]^2,$$

with respect to β and θ , where $x_{ij} = x_i/x_j$ for $x_j > x_i, j = 1, \dots, n_g$.

Introducing the notation

$$\bar{Y}_{n_g} = \frac{1}{n_g} \sum_{j=1}^{n_g} Y_n(x_j)$$

it is easily seen that for this fixed point x_i , the least squares estimates of β and θ are given by

$$\begin{aligned} \hat{\theta}(x_i) &= \frac{\sum_{j=i+1}^{n_g} Y_n(x_j)(1 - x_{ij}) - \bar{Y}_{n_g} \sum_{j=i+1}^{n_g} (1 - x_{ij})}{\sum_{j=i+1}^{n_g} (1 - x_{ij})^2 - \frac{1}{n_g} \left[\sum_{j=i+1}^{n_g} (1 - x_{ij}) \right]^2} \\ \hat{\beta}(x_i) &= \bar{Y}_{n_g} - \frac{1}{n_g} \hat{\theta}(x_i) \sum_{j=i+1}^{n_g} (1 - x_{ij}). \end{aligned}$$

These estimated values of θ and β are then used to find the fitted value of $Y(x_j)$, for any grid point x_j in the initial grid. Referring to (2.7), for the fixed point x_i , this fitted value

is given by

$$Y_{n,x_i}(x_j) = \hat{\beta}(x_i) + \hat{\theta}(x_i)(1 - x_{ij})I_{\{x_i < x_j\}},$$

for all the grid points x_j . The residual sum of squares resulting from estimating the unknown τ by the grid point x_i is then

$$\text{RSS}(x_i) = \sum_{j=1}^{n_g} [Y_n(x_j) - Y_{n,x_i}(x_j)]^2.$$

The estimator $\hat{\tau}_{\text{LS},\alpha_1,\alpha_2}$ of τ that we propose is now defined as the grid point $x_j \in I_{\alpha_1,\alpha_2}$ where $\text{RSS}(\cdot)$ attains its minimum value.

3. Data Analysis

We now analyze the data described in the introduction using model (2.1). We report results for all three estimation methods discussed in Section 2, and for all methods τ is estimated first and once an estimator $\hat{\tau}$ is obtained, the estimators of β and θ are obtained via maximum likelihood estimation.

Lymphoma Data: In their study, Matthews, Farewell and Pyke (1985) find a very small significance level for the null hypothesis of constant hazard rate using $\tau_0 = 10$ and $\tau_1 = 35$. We used the same $[\tau_0, \tau_1]$ interval for the estimation of τ for all three methods. The function $Y_n(t)$, estimating $Y(t)$, is shown in Figure 3. The estimated values for the (τ, β, θ) triplet by the least squares estimation (LSE) method was $(22.09, 0.03, -0.011)$. In the LSE method, we set $\alpha_1 = 0$ and $\alpha_2 = 0.15$ in our analysis, since the censoring

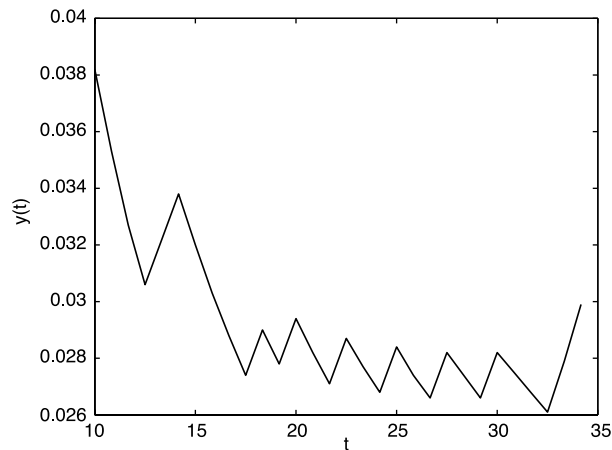


Figure 3. Estimated $Y(t)$ for the LSE estimator for the Lymphoma data.

Table 1. Results of MLE, LSE and CCH estimators for Stanford Heart Transplant data.

METHOD	$[\tau_0, \tau_1]$	τ	β	θ
MLE	[20, 120]	68.2352	0.0044	−0.0039
	[20, 200]	68.7058	0.0044	−0.0039
	[20, 500]	70.8235	0.0043	−0.0038
	[20, 1000]	71.8824	0.0042	−0.0037
LSE	[20, 120]	80.58813	0.0038	−0.0034
	[20, 200]	81.41171	0.0038	−0.0034
	[20, 500]	79.29412	0.0039	−0.0037
	[20, 1000]	71.88235	0.0042	−0.0038
CCH	[20, 120]	22.00000	0.0034	−0.0027
	[20, 200]	22.00000	0.0034	−0.0027

proportion is quite high in the right tail of the data. The maximum likelihood estimators (MLE) of the parameters were (18.065, 0.028, −0.004) and the Chang, Chen and Hsiung (CCH) estimator yielded (18.3, 0.029, −0.009). For the CCH estimator, $T = 50$ is used. All three estimation methods indicate a change point in the hazard function in the second half of the second year, and the MLE estimates a relatively smaller jump size than the other two.

Heart Transplant Data: Loader (1991) tested the null hypothesis of a constant hazard rate against the change point alternative and reported that the result is highly significant in favor of the alternative. In this example, in order to illustrate the sensitivity of the estimators to the choice of $[\tau_0, \tau_1]$, we selected four different search intervals as given in Table 1. The lower end of the search interval is taken to be $\tau_0 = 20$ and the cases $\tau_1 = 120, 200, 500, 1000$ are considered for the upper limit. In all cases 170 equally spaced grid points are considered. For all three methods the estimators of β and θ are close to each other, but the estimates of τ differ more. All methods capture the decreasing property of the hazard rate function by estimating a jump size of −0.003 to −0.004. The MLE and LSE methods for τ result in closer values as the search interval gets wider and in general LSE results in larger estimates for τ . The function $Y(y)$ in the LSE method is depicted in Figure 4 over the search interval [20, 120]. For the LSE estimator, the truncation proportion at the tails were taken to be $\alpha_1 = 0.03$ and $\alpha_2 = 0.05$, and other reasonable choices of these quantities did not make a difference in the estimation. We also note that for this particular data set, the CCH method sets the estimator of τ to the smallest grid point larger than the lower end point of the search interval, which was 22 in our choice. Therefore we report a single result for the CCH method.

4. Performances of the Estimators

We now investigate the performance of the least squares estimators for τ , β and θ via simulations. This performance will be compared with that of the maximum likelihood

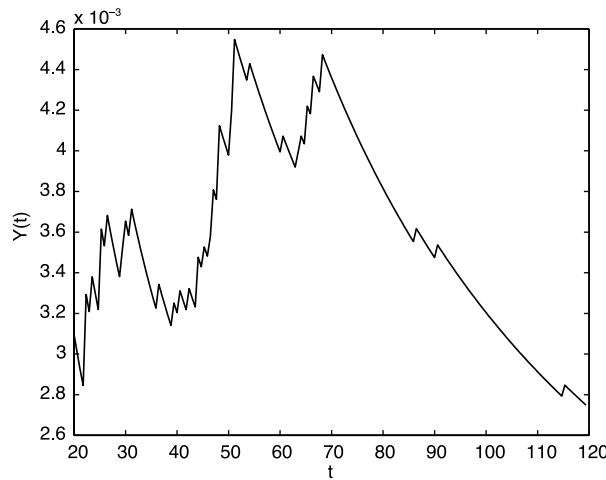


Figure 4. Estimated $Y(t)$ of the Stanford heart transplant data for the LSE estimator.

estimator described in Section 2.1 and the estimator studied by Chang, Chen and Hsiung (1994) from Section 2.2, referred to as the CCH estimator.

We simulated data from model (2.1), with $\tau = 1$, $\beta = 1$ and for the jump size θ we considered two values: $\theta = 1$ and $\theta = 0.7$, where the latter value of θ indicates a more difficult estimation task. For the censoring variable C we assumed a Uniform distribution on the interval $[0.4, 1]$, which results in a censoring proportion of about 20%.

Since each of the three procedures involves some choices of intervals and/or other parameters, we simulated data for a number of set-ups. The most detailed simulation study was done for the newly-proposed LSE estimator. We carried out an extensive simulation study for each of these estimation procedures, but only a subset of them are reported here for space considerations. The presented results are based on 1000 simulations for various sample sizes.

Tables 2, 3 and 4 summarize the results obtained for the MLE, CCH and the LSE respectively for $\theta = 1$. Each table reports on the estimated value and its associated bias and standard deviation for each of the three parameters.

From the above tables we can draw some conclusions on the performances of each of the three estimation procedures:

- * The choice of the interval $[\tau_0, \tau_1]$ as well as the choice of the upper limit T are quite important in the CCH procedure (Table 3 only reports on the choice of the upper limit, but see also Table 5 in this respect).
- * Both the MLE and the LSE estimator are rather insensitive to the choice of the interval $[\tau_0, \tau_1]$.
- * The choice of the 'left-out' proportions α_1 and α_2 are of minor importance in the LSE procedure. A small positive value for α_2 is recommendable.

Table 2. Performances of the maximum likelihood estimators for the true values of the parameters $\tau = 1$, $\beta = 1$ and $\theta = 1$.

[τ_0, τ_1]	n	estimation of τ			estimation of β			estimation of θ		
		$\hat{\tau}$	estimated		$\hat{\beta}$	estimated		$\hat{\theta}$	estimated	
			bias	stdev.		bias	stdev.		bias	stdev.
[0.25, 1.75]	50	1.089	0.089	0.360	0.980	-0.020	0.241	14.431	13.431	77.664
	100	1.067	0.067	0.293	0.986	-0.014	0.157	1.637	0.637	1.705
	400	1.025	0.025	0.113	0.998	-0.002	0.068	1.104	0.104	0.255
[0.5, 1.5]	50	1.062	0.062	0.245	0.978	-0.022	0.209	9.258	8.258	11.763
	100	1.060	0.060	0.203	0.989	-0.011	0.144	1.479	0.479	0.835
	400	1.024	0.024	0.096	0.998	-0.002	0.068	1.101	0.101	0.244
[0.75, 1.15]	50	1.004	0.004	0.105	0.976	-0.024	0.196	1.519	0.519	1.041
	100	1.009	0.009	0.094	0.984	-0.016	0.134	1.287	0.287	0.569
	400	1.014	0.014	0.057	0.996	-0.004	0.066	1.085	0.085	0.233

Based on the tables we can also draw the following overall conclusions regarding a comparison of the estimation procedures:

- ◇ The CCH method has overall the smallest standard deviation, followed by the MLE. The LSE has typically a somewhat larger variance, which is not surprising since it involves in fact two estimation steps, namely the nonparametric estimation of the cumulative hazard function, and the least squares fitting. For estimation of β and θ though, the variances of the three estimators are quite comparable.
- ◇ The least squares procedure clearly outperforms the two other estimation methods in terms of bias, having a far smaller bias for estimation of τ , β and θ . The biggest gain is noted for the estimation of β and θ .
- ◇ The maximum likelihood estimator for θ performs quite badly (recall that the estimator of θ is influenced by the estimators of τ and β).

Table 3. Performances of the CCH estimators for the true values of the parameters $\tau = 1$, $\beta = 1$ and $\theta = 1$.

[τ_0, τ_1]	T	n	estimation of τ			estimation of β			estimation of θ		
			$\hat{\tau}$	estimated		$\hat{\beta}$	estimated		$\hat{\theta}$	estimated	
				bias	stdev.		bias	stdev.		bias	stdev.
[0.25, 1.75]	2	50	1.167	0.167	0.317	1.015	0.015	0.193	2.371	1.371	7.751
		100	1.150	0.150	0.262	1.017	0.017	0.140	1.575	0.575	1.047
		400	1.045	0.045	0.119	1.006	0.006	0.070	1.102	0.102	0.270
[0.25, 1.75]	3	50	0.995	-0.005	0.358	0.973	-0.027	0.191	1.432	0.432	0.920
		100	1.079	0.079	0.290	1.000	0.000	0.135	1.326	0.326	0.623
		400	1.133	0.133	0.231	1.023	0.023	0.078	1.156	0.156	0.316
[0.25, 1.75]	4	50	0.675	-0.325	0.345	0.927	-0.073	0.221	0.937	-0.063	0.665
		100	0.846	-0.154	0.289	0.961	-0.039	0.153	0.990	-0.010	0.495
		400	0.966	-0.034	0.150	0.992	-0.008	0.067	1.007	0.007	0.253

Table 4. Performances of the least squares estimators for the true values of the parameters $\tau = 1$, $\beta = 1$ and $\theta = 1$.

$[\tau_0, \tau_1]$	(α_1, α_2)	n	estimation of τ			estimation of β			estimation of θ		
			$\hat{\tau}$	estimated		$\hat{\beta}$	estimated		$\hat{\theta}$	estimated	
				bias	stdev.		bias	stdev.		bias	stdev.
[0.25, 1.75]	(0,0)	50	1.020	0.020	0.405	0.997	-0.000	0.223	2.455	1.455	25.290
		100	0.995	-0.0053	0.351	1.003	0.003	0.221	1.229	0.229	1.260
		400	0.969	-0.031	0.155	1.004	0.004	0.073	0.961	-0.039	0.280
[0.25, 1.75]	(0,0.03)	50	0.990	0.010	0.400	0.996	-0.004	0.221	1.530	0.530	1.783
		100	0.990	-0.010	0.350	1.002	0.002	0.156	1.195	0.195	0.896
		400	0.969	-0.031	0.156	1.004	0.004	0.073	0.961	-0.039	0.280
[0.25, 1.75]	(0.05,0.05)	50	0.978	-0.022	0.395	0.996	-0.004	0.221	1.491	0.491	1.715
		100	0.984	-0.016	0.348	1.002	0.002	0.155	1.190	0.190	0.894
		400	0.969	-0.031	0.156	1.004	0.004	0.073	0.961	-0.039	0.280
[0.25, 1.75]	(0.03,0.07)	50	0.951	-0.049	0.382	0.994	-0.006	0.221	1.377	0.377	1.248
		100	0.958	-0.042	0.338	1.000	0.000	0.154	1.149	0.149	0.767
		400	0.969	-0.032	0.154	1.004	0.004	0.073	0.960	-0.040	0.279
[0.5, 1.5]	(0,0)	50	1.002	0.002	0.276	1.008	0.008	0.210	1.529	0.529	1.886
		100	0.999	-0.001	0.249	1.005	0.005	0.146	1.185	0.185	0.712
		400	0.975	-0.025	0.115	1.003	0.003	0.069	0.981	-0.019	0.266
[0.5, 1.5]	(0,0.03)	50	1.016	0.016	0.274	1.007	0.007	0.209	1.494	0.494	1.498
		100	0.999	-0.001	0.249	1.005	0.005	0.146	1.185	0.185	0.712
		400	0.975	-0.025	0.115	1.003	0.003	0.069	0.981	-0.019	0.266
[0.5, 1.5]	(0.05,0.05)	50	1.012	0.012	0.273	1.006	0.006	0.208	1.475	0.475	1.413
		100	0.999	-0.001	0.249	1.005	0.005	0.146	1.185	0.185	0.712
		400	0.975	-0.025	0.115	1.003	0.003	0.069	0.981	-0.019	0.266
[0.5, 1.5]	(0.03,0.07)	50	0.100	-0.000	0.270	1.005	0.005	0.208	1.429	0.429	1.248
		100	0.994	-0.006	0.247	1.005	0.005	0.144	1.177	0.177	0.692
		400	0.975	-0.025	0.115	1.003	0.003	0.069	0.981	-0.019	0.266
[0.75, 1.15]	(0,0)	50	0.987	-0.013	0.118	1.009	0.009	0.198	1.251	0.251	0.928
		100	0.986	-0.014	0.111	1.004	0.004	0.138	1.115	0.115	0.544
		400	0.989	-0.011	0.071	1.001	0.001	0.067	1.008	0.008	0.234
[0.75, 1.15]	(0,0.03)	50	0.987	-0.013	0.118	1.009	0.009	0.198	1.251	0.251	0.928
		100	0.986	-0.014	0.111	1.004	0.004	0.138	1.115	0.115	0.544
		400	0.989	-0.011	0.071	1.001	0.001	0.067	1.008	0.008	0.234
[0.75, 1.15]	(0.05,0.05)	50	0.987	-0.013	0.118	1.009	0.009	0.198	1.251	0.251	0.928
		100	0.986	-0.014	0.111	1.004	0.004	0.138	1.115	0.115	0.544
		400	0.989	-0.011	0.071	1.001	0.001	0.067	1.008	0.008	0.234
[0.75, 1.15]	(0.03,0.07)	50	0.987	-0.013	0.118	1.009	0.009	0.198	1.251	0.251	0.928
		100	0.986	-0.014	0.111	1.004	0.004	0.138	1.115	0.115	0.544
		400	0.989	-0.011	0.071	1.001	0.001	0.067	1.008	0.008	0.234

Table 5 reports similarly on the performances of the three procedures when $\theta = 0.7$, representing a more difficult estimation task. Similar conclusions as the ones drawn from Tables 2–4 can be formulated here, with an even more pronounced gain in terms of bias for the LSE estimator. For an additional evaluation we also produced kernel density estimates for the three parameters. In particular, we simulated 1000 samples of size

Table 5. Performances of the maximum likelihood estimators (MLE), CCH estimators (with upper limit $T = 4$) and least squares estimators (LSE) for the true values of the parameters $\tau = 1$, $\beta = 1$ and $\theta = 0.7$.

	$[\tau_0, \tau_1]$	n	estimation of τ			estimation of β			estimation of θ		
			$\hat{\tau}$	estimated		$\hat{\beta}$	estimated		$\hat{\theta}$	estimated	
				bias	stdev.		bias	stdev.		bias	stdev.
MLE	[0.25, 1.75]	50	1.031	0.031	0.366	0.937	-0.063	0.255	2.978	2.278	26.167
		100	1.003	0.003	0.284	0.928	-0.072	0.157	1.240	0.540	0.920
		400	0.874	-0.126	0.103	0.910	-0.090	0.073	0.864	0.164	0.173
	[0.50, 1.50]	50	1.012	0.012	0.242	0.927	-0.073	0.213	1.565	0.865	2.276
		100	0.980	-0.020	0.199	0.929	-0.071	0.146	1.118	0.418	0.590
		400	0.871	-0.129	0.090	0.914	-0.086	0.072	0.846	0.146	0.171
CCH	[0.25, 1.75]	50	0.749	-0.251	0.340	0.904	-0.096	0.209	0.905	0.205	0.606
		100	0.825	-0.175	0.270	0.914	-0.086	0.140	0.882	0.182	0.421
		400	0.886	-0.114	0.156	0.916	-0.084	0.074	0.850	0.150	0.185
	[0.50, 1.50]	50	0.856	-0.144	0.203	0.924	-0.076	0.193	0.958	0.258	0.588
		100	0.876	-0.124	0.152	0.921	-0.079	0.135	0.904	0.204	0.388
		400	0.876	-0.124	0.115	0.915	-0.085	0.072	0.844	0.144	0.174
LSE	[0.25, 1.75]	50	1.010	0.010	0.418	0.978	-0.022	0.216	1.243	0.543	2.012
		100	1.028	0.028	0.384	0.984	-0.016	0.154	0.965	0.265	0.830
		400	0.945	-0.055	0.187	0.974	-0.026	0.074	0.729	0.029	0.243
	[0.50, 1.50]	50	1.019	0.019	0.293	0.976	-0.024	0.208	1.131	0.431	1.113
		100	0.984	-0.016	0.252	0.972	-0.028	0.145	0.900	0.200	0.536
		400	0.928	-0.072	0.136	0.965	-0.035	0.072	0.744	0.044	0.227

$n = 300$ from model (2.1) with $\tau = \beta = 1$ and $\theta = 0.7$. The estimators for τ are searched for in the interval $[0.25, 1.75]$. The upper limit in the CCH method was chosen to be $T = 4.0$, and for the proportions in the least squares procedure we took $\alpha_1 = 0.0$ and $\alpha_2 = 0.03$. For each of the 1000 simulated samples, we obtained the estimated values for τ , β and θ . Then a kernel density estimate is obtained for each parameter estimator, using a Gaussian kernel and a bandwidth $h = 1.06S_n^{-1/5}$, where S^2 is the variance of the 1000 estimated parameter values. Figure 5 shows these estimated distributions for the estimates of τ , β and θ (top, middle and bottom panel respectively). The displayed density functions confirm the conclusions from the tables. Figure 5 clearly shows that the LSE estimator is far less biased, at a price of having a considerably bigger variance. The advantage of the LSE estimation procedure is most visible for the estimation of β and θ since there is a significant gain in terms of bias, whereas the variance only slightly increases.

Since our tables do not directly report on the mean squared error of the estimators, we plot the square root of the mean squared errors (rmse) as a function of the sample size for the three estimation procedures. These plots are based on the calculated mean squared errors based on 1000 simulations for sample sizes $n = 15, 25, 50, 75, 100, 150, 200, 300, 500$ and 800 . The parameter set-ups here were the same as those for Figure 5. The results for τ , β and θ are depicted in Figure 6 (respectively top, middle and bottom panel). It is observed from these plots that the maximum likelihood method performs better than the other two estimators for estimating τ for most of the sample sizes and for large samples,

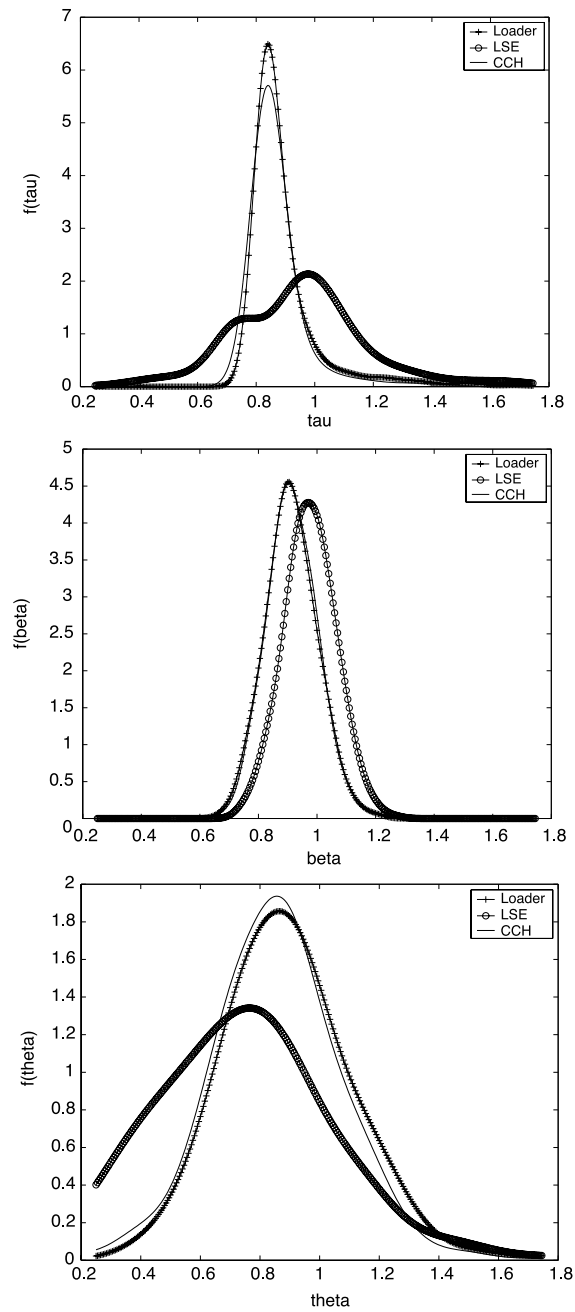


Figure 5. Density estimates for the distribution of $\hat{\tau}$ (top), $\hat{\beta}$ (middle) and $\hat{\theta}$ (bottom) for the three estimation procedures, based on 1000 samples of size $n = 300$. True values: $\tau = \beta = 1$ and $\theta = 0.7$.

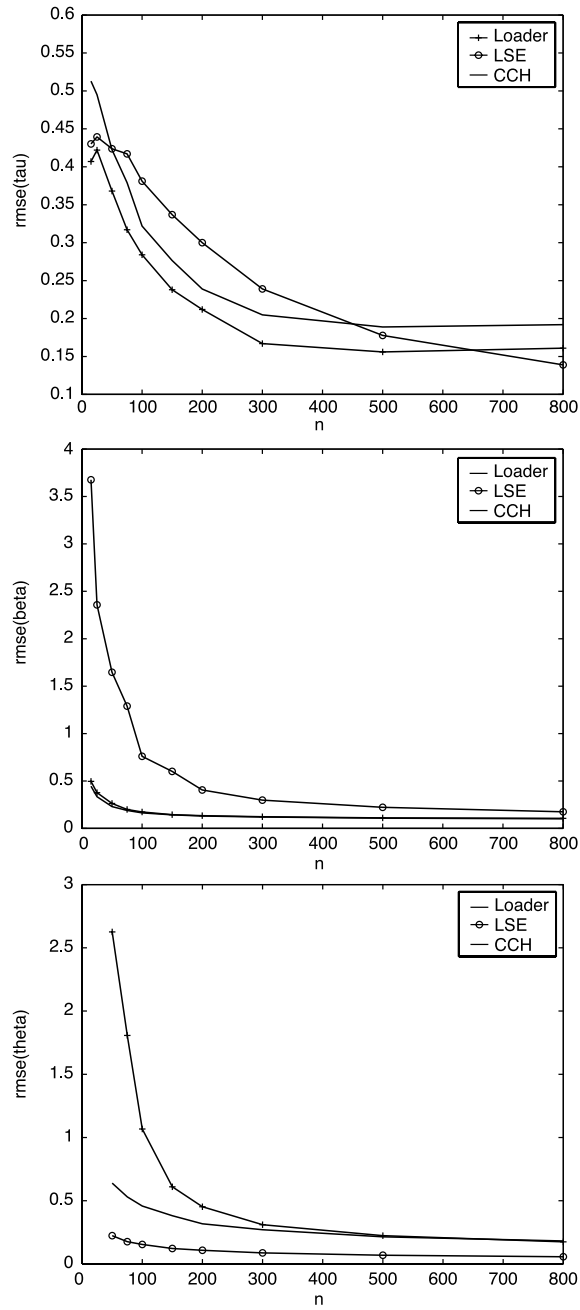


Figure 6. Root mean squared errors of the three estimators for τ (top), β (middle) and θ (bottom) as a function of sample size, based on 1000 simulations.

the rmse of the LSE decreases more rapidly than the other two. For the estimation of β , the CCH and the MLE perform almost similarly and better than the LSE, and the difference being maximum for smaller sample sizes. Finally, it is observed that the LSE performs notably better than the others for estimating θ , uniformly for all sample sizes.

Acknowledgments

The authors thank an Associate Editor and a Referee for their valuable remarks which led to an improvement of the paper.

Note

1. This research is partially supported by NATO Collaborative Research Grant CRG 950271, by 'Projet d'Actions de Recherche Concertées', No. 98/03-217, from the Belgian government, and by the IAP research network nr P5/24 of the Federal Office for Scientific, Technical and Cultural Affairs, Belgium.

References

- O. Aalen, "Nonparametric inference for a family of counting processes," *Ann. Statist.* vol. 6 pp. 701–726, 1978.
- I.-S. Chang, C.-H. Chen, and C. A. Hsiung, "Estimation in change-point hazard rate models with random censorship," in *Change-point problems, IMS Lecture Notes – Monograph Series* vol. 23 pp. 78–92, 1994.
- D. R. Cox and D. Oakes, *Analysis of Survival Data*, Chapman and Hall: London, 1984.
- R. Henderson, "A problem with the likelihood ratio test for a change-point hazard rate model," *Biometrika* vol. 77 pp. 835–843, 1990.
- C. R. Loader, "Inference for a hazard rate change point," *Biometrika* vol. 78 pp. 749–757, 1991.
- D. E. Matthews and V. T. Farewell, "On testing for a constant against a change-point alternative," *Biometrics* vol. 38 pp. 463–468, 1982.
- D. E. Matthews, V. T. Farewell, and R. Pyke, "Asymptotic score-statistic processes and tests for constant hazard rate against a change-point alternative," *Ann. Statist.* vol. 13 pp. 583–591, 1985.
- R. Müller and J. Halpern, "Regression with censored data," *Biometrika* vol. 69 pp. 521–531, 1982.
- H. G. Müller and J.-L. Wang, "Nonparametric analysis of changes in hazard rates for censored survival data: an alternative to change-point problems," *Biometrika* vol. 77 pp. 305–314, 1990.
- H.-G. Müller and J.-L. Wang, "Change-point models for hazard functions," in *Change-point problems, IMS Lecture Notes – Monograph Series* vol. 23 pp. 224–241, 1994.
- H. T. Nguyen, G. S. Rogers, and E. A. Walker, "Estimation in change-point hazard rate model," *Biometrika* vol. 71 pp. 299–304, 1984.
- Y.-C. Yao, "Maximum likelihood estimation in hazard rate models with a change-point," *Comm. Statist., Series A* vol. 15 pp. 2455–2466, 1986.